

## Rapid and Semi-Automated Analysis of 4D-STEM data via Unsupervised Learning

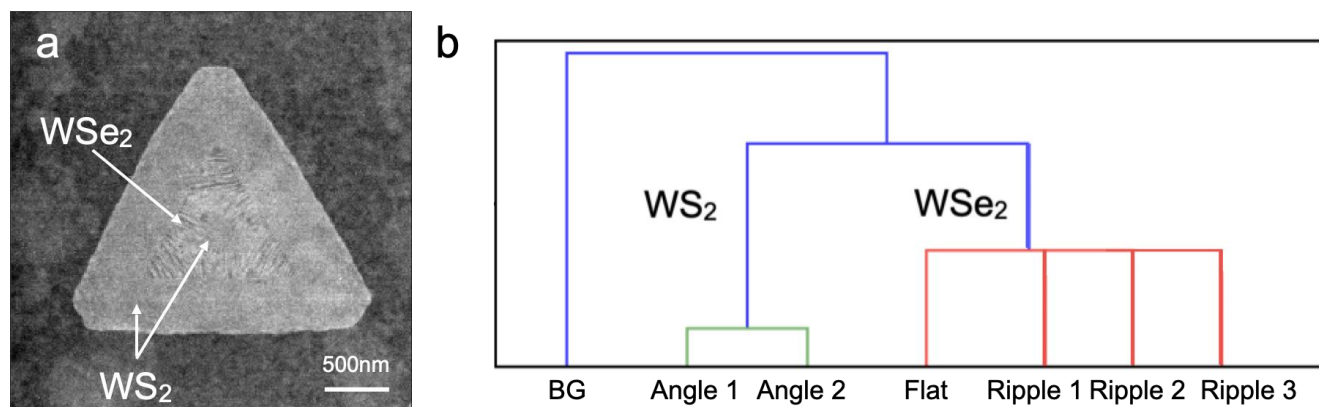
Chuqiao Shi<sup>1</sup>, Michael Cao<sup>1</sup>, David Muller<sup>2</sup> and Yimo Han<sup>1</sup>

<sup>1</sup>Rice University, United States, <sup>2</sup>School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA, Ithaca, New York, United States

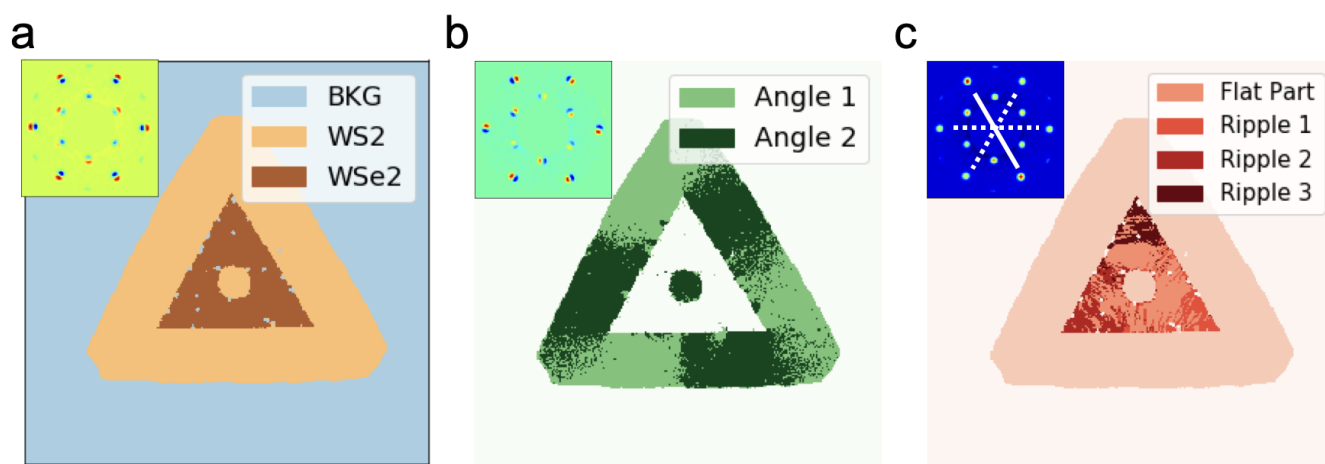
Four-dimensional scanning transmission electron microscopy (4D-STEM) is an advanced characterization technique, capable of mapping crystalline structures and deformations in novel nanomaterials with ultrahigh-resolution and precision<sup>1,2</sup>. However, it remains challenging to process and interpret large datasets generated from 4D-STEM. Machine learning has emerged as a promising method for microscopy due to its capability of analyzing complex patterns in large datasets<sup>3,4</sup>. Here, we developed a rapid and semi-automated approach based on unsupervised learning to process large 4D-STEM datasets.

Our method utilizes divisive hierarchical clustering to separate the 4D-STEM dataset into different regions according to similar lattice structures. This method provides the ability to distinguish different samples in one dataset and extract strain features in each sample without prior knowledge. Through this approach, lateral WS<sub>2</sub>-WSe<sub>2</sub> hetero-junctions were examined (Fig. 1a). Three rounds of clustering were applied on this 4D-STEM dataset (Fig. 1b). In the first and second round clustering, the junction sample is separated from the amorphous supporting film, and the two-component materials (WS<sub>2</sub> and WSe<sub>2</sub>) are distinguished (Fig. 2a). The difference between these two sample clusters is the lattice constant, which is reflected on the diffraction patterns (inset of Fig. 2a). In the third round of clustering, we applied unsupervised learning on the two sub-clusters of WS<sub>2</sub> and WSe<sub>2</sub>. Our method separates the WS<sub>2</sub> cluster into two more sub-clusters, where the real space map shows a rotational periodicity of the material with graded interfaces (Fig. 2b). The two clusters we achieved from our method in WS<sub>2</sub> distinguish the lattice rotational strain fields hidden in the single crystalline WS<sub>2</sub>. We also extracted out-of-plane ripples in different directions in the WSe<sub>2</sub> samples. The ripple causes the tilt of diffraction pattern, which is reflected by the intensity changes of different diffraction spots in the cluster center (inset of Fig. 2c). In addition to WS<sub>2</sub> and WSe<sub>2</sub>, our hierarchical clustering method can also be used to extract the strain features and deformation information in other nanomaterials from 4D-STEM datasets.

In conclusion, due to the importance of understanding the role of deformations in material properties and device performance, our method represents a crucial step towards a deeper and easier method for the analysis of deformations. This purely data-driven analysis uncovers different types of material deformations in the samples. In addition, our method is also crucial for overcoming the current barriers of 4D data analysis, generalizing this powerful technique for the study of atomic structures in a large scale and broadening material systems or other imaging techniques that generate large and multidimensional datasets, benefiting the materials science and microscopy community.



**Figure 1.** (a) ADF-STEM image of the WS<sub>2</sub>-WSe<sub>2</sub> junction. (b) Dendrogram of the unsupervised learning workflow.



**Figure 2.** (a) Real space label maps from the unsupervised learning results in the first and second round clustering. The difference of the cluster centers of WSe<sub>2</sub> and WS<sub>2</sub> diffraction pattern is shown on the top-left. (b) Real space label maps from the WS<sub>2</sub> sample. The difference of the cluster centers of the two rotations is shown. (c) Real space label maps from WSe<sub>2</sub> sample. Inset, the cluster center of the Ripple 2 part (red).

## References

- [1] Y. Han et al. *Nano Letters*, 18 (2018), p. 3746-3751.
- [2] Jiang, Y. *et al.* *Nature*, 559 (2018), 343–349.
- [3] Jesse, S. *et al.* *Sci Rep-uk* 6 (2016), 26348.
- [4] Lee, C.-H. *et al.* *Nano Lett* 20 (2020), 3369–3377.